AD-A070 024 CALIFORNIA UNIV BERKELEY STATISTICAL LAB F/G 12/1
CLUSTERING: REMINISCENCES OF SOME EPISODES IN MY RESEARCH ACTIV--ETC(U)
1979 J NEYMAN N00014-75-C-0159
UNCLASSIFIED CU-SL-79-03-ONR NL

END
DATE FILMED
7-79
DDC

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Statistical Laboratory
University of California
Berkeley, California 94720

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

(6) 3. REPORT TITLE
Clustering: reminiscences of some episodes in my research activity

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Scientific

(10) 5. AUTHOR(S) (First name, middle initial, last name)
Jerzy Neyman

(12) 23p.

(11) 6. REPORT DATE
1979 (15)

7a. TOTAL NO. OF PAGES
20

7b. NO. OF REFS
5

8a. CONTRACT OR GRANT NO.
ONR N000 14-75-C-0159,
b. PROJECT NO
c. DAAG29-76-G-0167
d.

9a. ORIGINATOR'S REPORT NUMBER(S)
ONR 79-03

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
(14) CU-SL-79-03-ONR

10. DISTRIBUTION STATEMENT
This document has been approved for public release; its distribution is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Office of Naval Research
Washington, D.C. 20014

13. ABSTRACT

The ideas of a stochastic process of clustering came to the author's attention from Dr. Geoffrey Beall, an entomologist interested in the distribution of larvae over an experimental field. Larvae are born from eggs deposited by moths, not singly, but in "egg-masses." After hatching, larvae begin to crawl in search of food. Later, a "general census" of larvae is performed. The r.v. of interest X = no. of larvae counted in a unit area plot in the field. Conceptual elements: cluster centers (= egg-masses), cluster size (= no. of larvae from a single egg-mass), dispersal of cluster members. Over the four decades since the publication of the theory relating to larvae, essentially the same mechanism of clustering was found to underly many diverse natural phenomena: clustering of galaxies, population dynamics, epidemics and effects of irradiation of living cells.

DD FORM 1 NOV 65 1473

Unclassified

Security Classification

333 400 LB

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Cluster center						
Cluster size						
Dispersal of cluster						

CLUSTERING:
REMINISCENCES OF SOME EPISODES
IN MY RESEARCH ACTIVITY*

Jerzy Neyman
Statistical Laboratory
University of California, Berkeley 94720

1. Introduction. It is a pleasure to be able to deliver this Second Pfizer Annual Colloquium. In selecting its subject I thought of work in our Berkeley Stat. Lab. relating to pharmacology. However, as of now, this work is not sufficiently advanced to be reported on this important occasion.

The subject I selected covers a long series of interconnected studies in several substantive domains, all of them reflecting the inspiration I received in the late 1930's from Dr. Geoffrey Beall, then of the Dominion Entomological Experimental Station, Chatham, Ontario. Regretfully, I have no references to Dr. Beall's publications.

It happened that, while Dr. Beall's preoccupation was with a special kind of entomological experiment, the idea for which I feel indebted to him, that of the phenomenon of clustering, proved to be very relevent in the following diverse domains: (i) in the study of spatial distributions of galaxies, (ii) in

---

*Approximate text of the presentation at The Pfizer Colloquium at the Department of Statistics, The University of Connecticut, April 30, 1979.

population dynamics, (iii) in the theory of epidemics, and (iv) in the study of radiation carcinogenesis.

While this presentation is intended to reflect my personal experiences, it may also be considered as a contribution to the history of a concept, the concept of "clustering." This historical sketch covers four decades. The unavoidable consequence is that most of the developments described are "symbolized" rather than studied in depth. Still, my hope is that the evolution of the original simple concept will be found intelligible.

The plan of the present paper is as follows. Section 2 outlines the original problem of Dr. Beall, concerned with counts of larvae in plots of an experimental field. The corresponding stochastic process may be labeled that of a "single clustering."

Section 3 is also concerned with the "single clustering" process. However, the substantive domain is very different: distribution of galaxies in space.

The subjects of all the subsequent sections are concerned with sequences of consecutive clusterings, that is, of the process of clustering of clusters. The natural phenomena studied include population dynamics (Section 4), radiation carcinogenesis (Section 5), and theory of epidemics, first "outdated" (Section 6) and later "modernized" (Section 7). Section 8: Concluding remarks.

2. Single Clustering: Counts of Larvae in Plots of an Experimental Field. Consider a large, reasonably uniform, ex-

perimental field divided into a number of unit area plots. Consider one of these plots, which it will be convenient to describe as "target." At a particular period during the summer, moths are flying over the field and, from time to time, deposit their eggs. These eggs are not deposited singly but in "masses," each composed of a large number of eggs. In due course, these eggs produce larvae which begin to crawl in search of food. After a certain period of time, when larvae are somewhat larger and convenient to count, a general census is performed and our interest is concerned with the number, say X, of larvae counted in the target plot.

The concept of clustering is connected with the fact that the larvae cannot travel fast. The possibility of one of them being found in the target plot depends on the distance of the egg-mass (= "cluster center") from which the given larva emerged. Thus, the conceptual counterpart of the empirical phenomenon relating to a single "target" plot must involve a bigger area, the "area of accessibility" surrounding the target.

The whole mechanism of clustering involves, then, the following concepts: (i) the distribution of egg-masses (= "cluster centers") over the field, (ii) the number of larvae from a single mass surviving up to the census (= number of cluster members, or "size" of the cluster), (iii) the mechanism of "dispersal" of cluster members and the implied "area of accessibility."

Naturally, the details of these three conceptual elements must vary from one empirical domain of study to the next. Figures 1 and 2, both representing the details of the original publication of 1939, are intended to "symbolize" the contemporary thinking. It will be seen that at the time when the paper was written, in 1936 or 1937, Dr. Beall and I did not dream that the mechanism of clustering could be relevent to the understanding of phenomena of clustering of galaxies, etc. The then contemplated applications were "entomology" and "bacteriology."

ON A NEW CLASS OF "CONTAGIOUS" DISTRIBUTIONS, APPLICABLE IN ENTOMOLOGY AND BACTERIOLOGY

By J. Neyman

CONTENTS

Figure 1. A detail of the title page of the original paper of 1939. Ann. Math. Stat., Vol. 10 (1939), pp. 35-57.

TABLE I

Distribution of European corn borers in 120 groups of 8 hills each, (data provided by Dr. Beall), fitted by Poisson Law and by type A Law with two parameters

No. of borers	Frequency		
	Exp. P. L.	Observed	Exp. T. A.
0	5.0	24	22.6
1	16.0	16	16.7
2	25.3	16	18.3
3	26.7	18	16.4
4	21.1	15	13.4
5	13.4	9	10.3
6	7.1	6	7.5
7	3.2	5	5.2
8	1.3	3	3.5
9	.4	4	2.3
10	.1	3	1.5
11		0	
12		1	
Beyond		—	2.3
m_1	—	—	2.178
m_2	—	—	1.454
P_{χ^2}	.000,000		.95

TABLE II

Distribution of yeast cells in 400 squares of haemacytometer observed by "Student" (1907), fitted by Poisson Law and by type A Law with two parameters

No. of cells	Frequency		
	Exp. P. L.	Observed	Exp. T. A.
0	202	213	214.8
1	138	128	121.3
2	47	37	45.7
3	11	18	13.7
4		3	3.6
5		1	.8
Beyond	2	—	.1
m_1	—	—	3.605
m_2	—	—	.189
P_{χ^2}	$> .02$		$> .1$

Figure 2. Two tables published at the end of the original paper of 1939.

3. Clustering of Galaxies. As is generally known, California is the "land of big telescopes." Having lived in Berkeley since August, 1938, it was unavoidable for me to become exposed to statistical problems of astronomy and, more specifically, of "extragalactic astronomy" or cosmology. In particular, I must record the inspiring influence of two "red-blooded" astronomers, N.U. Mayall and C.D. Shane, at the time both at the Lick Observatory, of which Shane was the director. The principal subject of our studies was the question whether, by and large, the distribution of galaxies in space is clustered, or, as was broadly believed, are the galaxies distributed in space singly, perhaps approaching a Poisson process. Beginning with 1952, there resulted a substantial sequence of publications, frequently co-authored by astronomers. These are exemplified by the following references:

J. Neyman and E.L. Scott, "A theory of spatial distribution of galaxies," Astrophysical Journ., Vol. 116 (1952), pp. 144-163.

J. Neyman, C.D. Shane and E.L. Scott, "On the spatial distribution of galaxies: a specific model," Astrophysical Journ., Vol. 117 (1953) pp. 92-133.

E.L. Scott, "The brightest galaxy in a cluster as a distance indicator," Astronomical Journ., Vol. 52 (1957), pp. 278-295.

J. Neyman, "Sur la théorie probabiliste des amas de galaxies et la vérification de l'hypothèse de l'exansion de l'univers," Annales de l'Institut Henri Poincaré, Vol. 14 (1955), pp. 201-244.

J.L. Lovasich, N.U. Mayall, J. Neyman, and E.L. Scott, "The expansion of clusters of galaxies," Proc. Fourth Berkelely Symp. on Math. Stat. and Prob., (J. Neyman, ed.), Univ. of Calif. Press, Berkeley and Los Angeles, Vol. 3 (1961), pp. 187-227.

It will be realized that, compared to clustering of larvae in a field, the cosmological aspect of the phenomenon of clustering is immeasureably more complex. One reason is the impossibility of approaching the cluster and counting its members! All the astronomer can do is to look at photographs of the sky, count on them the images of galaxies, study images of single galaxies, and also the spectra of the light they emit. The ingenuity the astronomers exhibit is really remarkable. Also, there is here a most encouraging east/west intellectual cooperation. This I had the pleasure of describing in my latest publication dealing with cosmology. The reference is: "Reminiscences of a Revolutionary Period in Cosmology," Problems of Physics and Evolution of the Universe (Festschrift for V.A. Ambartsumian), L.A. Mirzovan, ed., Publ. House of the Armenian Acad. of Sciences, Yerevan (1978), pp. 243-249.

4. Population Dynamics: Sequence of Clustering of Clusters, of Clusters, etc. Here the most relevent publication, co-authored with E.L. Scott, has the title: "On a mathematical theory of populations conceived as conglomeration of clusters." It appeared in 1957 in Vol. XXII of Proc. Cold Spring Harbor Symposia on Quantitative Biology, pp. 109-120.

The problem studied can be summarized as follows. Consider an infinite plane H representing the "habitat" and let $R_1, R_2, \ldots, R_s$ be any arbitrarily selected non-overlapping

regions in H. Also, let $m_1, m_2, \ldots, m_s$ be s arbitary non-negative integer numbers. The characterization of the distribution of a population inhabiting H is understood to mean a rule for determining the probability that, simultaneously, the numbers of the population members in R_1 will be exactly m_1, that the number of them located in R_2 will be exactly m_2, etc.

In other words, if X_i stands for the number of population members in R_i, the problem was to deduce the formula for the probability generating function of the random variables X_1, $X_2, \ldots, X_s$.

The mathematical assumptions used in the work are reducible to a repetition of the process of single clustering. A litter, having one or more members, is born at a point (= cluster center) in H. The members of the litter (cluster) disperse and gradually die out. Before dying, some of the litter members produce their own litters of progeny, etc. The particular object of study is the joint distribution of two successive generations of the population, the paternal and the filial. Also, some asymptotic results are obtained.

While the distribution of a species over the habitat appears as a domain very different from that of the distribution of galaxies in space, there are some important analogies. In either case no definitive empirical verification of the hypo-

thetical details is possible. All one can do is to perform a Monte Carlo simulation and to compare the results with such fragmentary observations as may be possible to accumulate.

5. Radiation Carcinogenesis. Out of the problems we studied in our Berkeley Stat. Lab., the chance mechanisms govern-ign carcinogenesis, particularly the radiation carcinogenesis, may well be the most difficult and, perhaps,the most important. I try to describe it ahead of epidemics for the reason that, at the present moment there is something like a "lull" in our efforts.

Obviously, in order to achieve some significant results leading to the understanding of the chance mechanism, or mechanisms, in living cells or tissues, a statistician must depend upon interested cooperation of an experimenting biologist. It happens that, at this moment, we lack the necessary contacts. On the other hand, as will be described in Sections 6 and 7, our studies of the mechanisms of epidemics develop at a reasonable rate.

Because the phenomenon of radiation carcinogenesis appears very distant from that of the distribution of larvae and, certainly, from cosmology, one is likely to believe that the underlying chance mechanisms could have nothing in common. Yet, closer examination indicates the contrary. Such differences as exist are differences of complexity.

The problems studied and the results obtained are summarized in a relatively recent paper written jointly with Prem S. Puri.

This paper being just a summary report, the best way to "symbolize" briefly the essence of our efforts seems to be through extensive quotes from our paper, including pictorial illustration.

Figure 3 reproduces the title page of our article. The article is concerned with the chance mechanism of damage to living cells caused by irradiation. One aspect of the "damage" may be the cell becoming cancerous. In parallel with the "damage" we consider the mechanism of possible "repair." Figures 4 and 6 illustrate the observable phenomena that our "structural model" is intended to explain. These phenomena include the difference between the so-called "high" and "low" linear energy transfers (LET.)

The reader will notice that the role of "egg-masses" in Dr. Beall's studies is now played by "primary" particles of irradiation. Experiments are possible to estimate their temporal distribution. This contrasts with the practical impossibility of counting the egg-masses. On the other hand, while in Dr. Beall's situation basic observables are counts of larvae in the test plots, in the problem of carcinogenesis the corresponding entities, namely the "secondary particles" and their "hits" in the targets, are impossible to count. The exception is the possibility of concluding that the number of "unrepaired" hits is zero, etc.

One easily illustrated common element is the "area" (or "volume") of accessibility.

Reprinted from
Proc. Natl. Acad. Sci. USA
Vol. 73, No. 10, pp. 3860–3863, October 1976
Statistics

A structural model of radiation effects in living cells

(mechanism of clustering/low linear energy transfer/high linear energy transfer/dose–response/dose rate)

JERZY NEYMAN* AND PREM S. PURI†

* Statistical Laboratory, University of California, Berkeley, Calif. 94720; and † Department of Statistics, Purdue University, West Lafayette, Indiana 47907

Communicated by Jerzy Neyman, July 12, 1976

ABSTRACT The chance mechanism of cell damage and of repair in the course of irradiation involves two details familiar to biologists that thus far seem to have been overlooked in mathematical treatment. One of these details is that, generally, the passage of a single "primary" radiation particle generates a "cluster" of secondaries which can produce "hits" that damage the living cell. With high linear energy transfer, each cluster contains very many secondary particles. With low linear energy transfer, the number of secondaries per cluster is generally small. The second overlooked detail of the chance mechanism is concerned with what may be called the time scales of radiation damage and of the subsequent repair. The generation of a cluster of secondary particles and the possible hits occur so rapidly that, for all practical purposes, they may be considered as occurring instantly. On the other hand, the subsequent changes in the damaged cells appear to require measurable amounts of time. The constructed stochastic model embodies these details, the clustering of secondary particles and the time scale difference. The results explain certain details of observed phenomena.

We use the terms *structural* or *stochastic* models of a phenomenon to designate a chance mechanism defined in terms of some hypothetical entities having some specified hypothetical properties, a mechanism the operation of which is expected to mimic the phenomenon studied. A stochastic model, a concept akin to Borel's idea of the principal problem of mathematical statistics (1) is contrasted with the brief term *model* now frequently encountered in statistical literature. This brief term is used to designate a more or less complicated formula, invented to fit the observations without any consideration of the mechanism that might have produced them. For this kind of "model" our preferred term is *interpolatory procedure*. Undoubtedly, such procedures are useful and, in fact, they appear unavoidable when an effort is made to adjust the details of a stochastic model to fit the observations.

The ultimate goal of the present study is a stochastic model of phenomena developing in irradiated experimental animals. However, the present paper is limited to irradiation effects on cells of some homogeneous tissue. The literature on this subject is quite rich. For example, see two recent papers by Payne and Garrett (2, 3). However, there appear to be many points of vagueness interestingly discussed by Mole (4).

The plan of the paper is as follows. First, we outline certain frequency findings related to experimental animals. At least some of these findings must be credited to Upton *et al.* (5). Others are taken from Totter (6). It is these findings that our model is intended to explain. The phenomena of interest have two aspects. One is biological and depends upon properties of living cells. The other aspect is physical, depending upon properties of radiation of one kind or another, e.g., high linear energy transfer (LET) and low LET. After illustrating these two aspects taken separately and in combination, we offer our model. Compared to its ancestors which came to our attention, the proposed model takes into account the striking difference between what may be called the time scales appropriate to biological and to physical aspects of the phenomena.

Abbreviations: LET, linear energy transfer; pgf, probability generating function.

Empirical findings

The empirical findings which stimulated the present paper are illustrated in the following two diagrams. The first, our Fig. 1, illustrates the so called "dose-rate effect" of gamma radiation (low LET) on the induction of a particular leukemia in mice (5). The point is that the same total amount of irradiation can be administered uniformly either over a long or over a relatively short period of time. In the former case, we speak of "low" dose rate and in the second of "high" dose rate. The graphs in Fig. 1 indicate that, with a "high" dose rate (the upper curve), a substantially higher percentage of irradiated mice acquire leukemia than with the low dose rate. Similar results were found for other cancers.

Fig. 1 illustrates also another phenomenon. This is that, perhaps unexpectedly, the effectiveness of gamma radiation administered at a high dose rate in inducing leukemia is not a monotone function of the dose. The observed frequency of the particular leukemia begins by increasing with the dose, reaches a maximum, and then decreases.

Fig. 2 is reproduced from an article by Totter (6). It compares the life-shortening effects of gamma rays and of neutrons, both administered at various dose rates and at various doses. The several curves relating to gamma rays exhibit strong dose-rate effects. Also, some of them suggest the presence of the maximum of dose effectiveness clearly shown in Fig. 1. Turning to neutrons, we see the same indication of a maximum of dose effect but, curiously, no noticeable dose-rate effect to the left of the point of maximum.

The "life shortening" indicated in Fig. 2 seems to have been measured by comparing the median life span of irradiated mice with that of the controls.

The physical properties of the various kinds of radiation, the properties that are particularly relevant to our work, are illustrated in Figs. 3 and 4. Fig. 3 represents a photograph of a cloud chamber exposed to a certain kind of irradiation. We are grateful to Alexander Crendon of the Donner Laboratory, University of California at Berkeley, for letting us use this photograph. The particularly relevant detail of Fig. 3 is the presence of crisscrossing lines, a few rather broad and many very thin. The lines mark the tracks of certain particles. They are composed of minute droplets formed about ions generated by the particles. Where the visible line is broad, the passage of the particle is accompanied by the appearance of many ions which travel to considerable distances away from the particle's track. Otherwise, there are only relatively few ions.

An important detail to be added to the facts illustrated in Fig. 3 is that the particles in question travel at enormous speeds, so that they cross a cell within a minute fraction of a second.

Figure 3

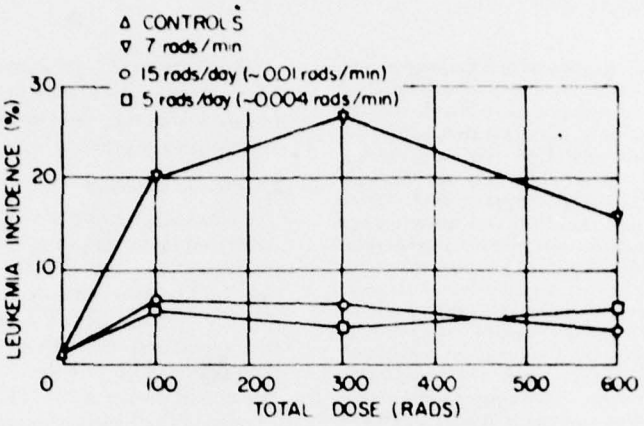

Figure 4. Incidence of myeloid leukemia in relation to dose and dose rate of gamma radiation. One rad = 0.01 J/kg.

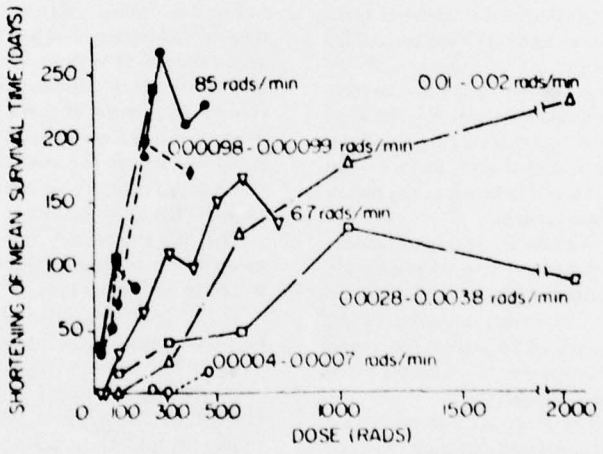

Figure 5. Life shortening in female mice as influenced by dose rate of gamma rays and neutrons. Open symbols represent gamma rays; filled symbols, neutrons.

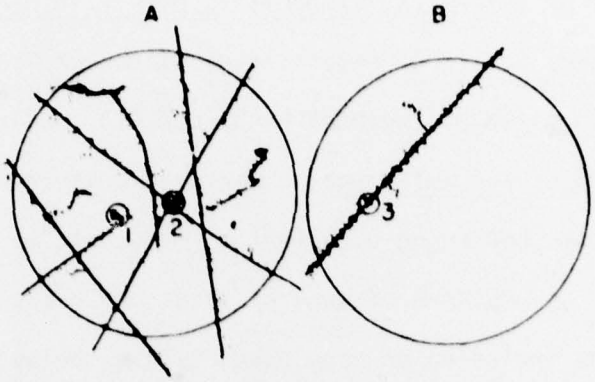

Figure 6. Schematic representation of ionization distribution in small volumes within a cell, irradiated with equal doses of x-rays (A) or α-radiation (B).

Figure 7. Photograph of a cloud chamber.

6. Theory of Epidemics (Outdated). Our Stat. Lab's effort at a theory of epidemics was published in 1964 in a paper co-authored by E.L. Scott and myself. The reference is "A Stochastic Model of Epidemics" (Stochastic Models in Medicine and Biology, J. Gurland, ed., The University of Wisconsin Press, 1964). The paper was inspired by the book by Norman T.J. Baily The Mathematical Theory of Epidemics that summarized quite a few earlier investigations, beginning with that of McKendric of 1926. One of the basic assumptions of many of these works was that, given a population including so many susceptibles the appearance of a single infectious creates a probability of contracting the disease that is the same for each of the susceptibles. As Bailey points out, any assumption of this kind may be realistic for a dormitory of a boarding school but not for a city and certainly not for a country. The stochastic model we produced is explicit in recognizing the lack of uniformity of the habitat. For a time, the model appeared reasonably realistic. However, during the winter quarter of 1978, there came the awareness of an important lack of realism. The ideas underlying an effort at modernization are described in the next section. The basic assumptions of the "outdated" theory are as follows.

(i) Number Infected by a Single Infectious. Consider an infinite plane H described as the habitat. It is assumed that to each point in H with coordinates $u=(u_1, u_2)$ there corresponds

a random variable $\nu(u)$ representing the number of susceptibles who would be infected if at that point there was a single infectious. While the actual distribution of $\nu(u)$ is left unspecified, it is assumed that the variables $\nu(u)$ corresponding to different points in the habitat are mutually independent. In fact, it is assumed that the variables $\nu(u)$ are independent of all other random variables of the system and that $p\{\nu(u)=0\}<1$.

(ii) Dispersal of Infected. It was assumed that during the "latent period" T (the same for all infected) the individuals infected at u travel independently from each other. Furthermore, it was assumed that to each point u in the habitat there corresponds a function $f(x|u)$ representing the probability density of the location $x=(x_1, x_2)$ where an individual infected at u becomes infectious. Except for certain conditions of regularity, the function $f(x|u)$, the "dispersal function" is left unspecified.

(iii) Immigration. The term "immigrants" is used to describe real infectious immigrants and also local inhabitants who become infectious "spontaneously" perhaps due to mutations of bacteria in their bodies. It is postulated that the appearance of an infectious "immigrant" is governed by a density function $\lambda(u)$ defined over the whole habitat and subject to certain conditions of regularity.

(iv) Discrete Generations. It was assumed that the duration

of infectiousness is zero and that occurrences of infection, all over the habitat occur simultaneously. In consequence, the development of an epidemic is divided into discrete generations.

The relevant mathematics involves the following two concepts. (i) The random variable $\nu(u)$, the number infected by a single infectious at a specified point $u=(u_1, u_2)$ in the habitat, and (ii) the random variable, say $\nu(X|u)$, the number infected somewhere in the habitat (X is a random variable) by a single individual of the earlier generation of the epidemic who became infected at a specified point u in the habitat.

The specification of a particular kind of epidemic, say of polio, depended on two families of functions, $\nu(u)$ (= the sizes of "clusters" centered at u) and the dispersal function $f(x|u)$, both subject to certain conditions of regularity.

The subjects of study included the possibility that an epidemic started by a single infectious might get "out of hand," as was once the case of a polio epidemic. Among the particular cases considered there was the possibility that a region R, marked by highly hygienic conditions, will escape the outbreak of a substantial epidemic, while in the rest of the habitat that same epidemic will get "out of hand." Two particular theorems, which in private conversations were called "democracy theorems" indicated that efforts at the establishment of such especially "healthy" regions would be futile.

7. Theory of Epidemics (Modernized). During the winter quarter of 1978, discussion of the theory of epidemics just described benefited by the participation of Mrs. Florence Morrison of the California State Department of Health. Also, we had several other rather interested and active members of the group of whom I shall mention two visitors from abroad, Dr. S. Kwesi Odoom, a Fulbright Fellow from Uganda, and Dr. Luis R. Perrichi from Venezuela.

Mrs. Morrison's most valuable contribution was the remark that the real habitats represented by entire countries are much more heterogeneous that the older theory presupposed. With reference to an epidemic of a communicable disease, Mrs. Morrison contended (and everyone agreed) that real habitats, such as the state of California, are stratified according to socio-economic status of the population. This stratification influences the development of an epidemic. At the very least three categories of locations have to be considered, depending on the income of the inhabitants: "high," "middle," and "low" (say slums, which is the term I used). There was the concensus of opinion that the number infected by a single infectious depends not only on the region, say R_i, in which the infection takes place, but also on the region, say R_j, where the infecting individual lives. For example, if an inhabitant of slums suddenly becomes infectious in the locality he inhabits, he is likely to infect many more people around him, than would the visiting

inhabitant of a high income region, etc.

The discussion that followed resulted in a somewhat unusual "take home exam" I formulated last year, see next page.

The result of the exam proved quite interesting to several participants in the discussion, including myself. While the subdivision of California into only three different socio-economic regions represented an obvious over-simplification, the results obtained appear instructive and there are plans afoot to produce a paper for publication.

8. Concluding Remarks. As mentioned at the outset, in selecting the subject of this Colloquium presentation, I had in mind to illustrate the phenomenon of evolution of an idea. The idea of the mechanism of "clustering" does not represent anything unique. Many other fruitful ideas also evolve. Otherwise, they would hardly be considered "fruitful." Ordinarily, the process of substantial evolution of a simple idea takes quite some time, much longer than the time we ordinarily spend in learning our contemporary state of that evolving idea. Thus, the phenomenon of the evolution escapes our attention. Yet, it seems interesting.

TAKE HOME FINAL EXAM

Due for delivery and discussion on Thursday, March 23, 12:30-3:30 pm, in Room 72 Evans.

Your presence at the discussion IS NECESSARY.

Instructions: Write clearly and tidily. Use ink or an intense black pencil.

* * * * * * *

Problem 1. State the basic assumptions of the theory presented in lectures and describe the principal results (e.g. What are the "democracy" theorems?). Use your own words. Do not copy from the published paper.

Problem 2. Criticize the basic assumptions of the theory even with reference to communicable diseases spread through personal contacts between infectious and susceptibles, like coughing. How should the basic assumptions be modified to make the theory more realistic?

Problem 3. Use computer facilities to simulate the development of an epidemic in conditions slightly more realistic than described in the early part of the course.

Consider a habitat composed of three regions R_1, R_2 and R_3:

R_1: high income region, with sparce population and facilities for travel.

R_2: middle income region (typified by Berkeley), with substantially denser population and with reasonable facilities for travel.

R_3: low income region, with very dense population and very limited travel opportunities.

Assume that each of the three regions is uniform in all respects and assign to them numerical values of the various relevant parameters (e.g. of probabilities that an individual inhabiting R_i will become infectious in R_i, etc.). The values assigned should be consistent with your intuition, but different from those of your colleagues (consult with them!). Next, consider an epidemic initiated by a single individual who became infectious in one of the three regions. Then, use the Monte Carlo simulation technique to generate 100 epidemics and calculate the mean number of cases in the successive generations of the epidemic and the mean total size of the epidemic. What about the "democratic" theorems?

Good Luck!

ACKNOWLEDGEMENTS

This paper was prepared using the facilities of the Statistical Laboratory with partial support from the Office of Naval Research (ONR N00014 75 C 0159), the Department of the Army (Grant DA AG 29 76 G 0167), and the National Institute of Environmental Health Sciences (2 R01 ES01299-16). This support is gratefully acknowledged. The opinions expressed are those of the author.